DOCUMENT RESUME

ED 080 645 UD 013 765

AUTHOR Jensen, Arthur R.
TITLE The Effect of Race of Examiner on the Mental Test Scores of White and Black Pupils.
PUB DATE 73
NOTE 32p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Ability; Caucasians; *Caucasian Students; Cognitive Tests; *Elementary Schools; *Examiners; Intelligence Tests; Negroes; *Negro Students; Nonverbal Tests; Perception Tests; Perceptual Motor Coordination; Test Results; Verbal Tests

ABSTRACT

An entire elementary school system with 60 percent white and 40 percent black children was given several ability tests administered by 12 white and eight black examiners. The tests measured verbal and nonverbal IQ, perceptual-motor cognitive development, "speed and persistence" under neutral and motivating instructions, listening attention, and short-term rote memory for numbers. With the exception of the "speed and persistence" test, on which the white examiners yielded higher mean scores than black examiners for both white and black pupils, the results for the cognitive ability tests showed that the race of the examiner had unsystematic and negligible effects in the testing of white and black pupils. These conclusions apply to the IQ test results of both group-administered as well as individually-administered tests. (Author)

FILMED FROM BEST AVAILABLE COPY

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 080645

The Effect of Race of Examiner on the Mental Test Scores of White and Black Pupils

Arthur R. Jensen

University of California, Berkeley

ABSTRACT

An entire elementary school system with 60% white and 40% black pupils was given several ability tests administered by 12 white and 8 black examiners (Es). The tests measured verbal and nonverbal IQ, perceptual-motor cognitive development, "speed and persistence" under neutral and motivating instructions, listening-attention, and short-term rote memory for numbers. With the exception of the "speed and persistence" test, on which white Es yielded higher mean scores than black Es for both white and black pupils, the results for the cognitive ability tests showed that the race of the E had unsystematic and negligible effects in the testing of white and black pupils. This conclusion applies to the IQ test results of both group-administered and individually-administered tests.

UD 013765

The Effect of Race of Examiner on the Mental Test Scores of White and Black Pupils[1]

Arthur R. Jensen

University of California, Berkeley

How often is it said that the race of the examiner is an important variable in the ability testing of ethnic minority children, or that black children obtain lower scores when tested by a white examiner? Sattler (1970), in a review of the research on this question, remarked: "In spite of the paucity of research concerning the effects of differences in racial status as a variable which affects the examiner-examinee relationship, numerous writers have either concluded or suggested that this variable may play an important role in the intelligence test situation." (p. 144). And in this and another review (Sattler & Theye, 1967, p. 353) Sattler cites a dozen references holding this belief, including books and articles by such noted psychologists as Anastasi, Hilgard, Klineberg, Pettigrew, Pressey, and Strong.

Though the speculative claims of race of examiner (E) effects in intelligence testing are frequent in the literature of race differences, the total empirical research on the subject, eleven studies and a reanalysis of one of these, altogether constitute a rather unimpressive body of evidence. They are here briefly summarized chronologically.

Canady (1936): On the first administration of 1916 Stanford-Binet,

Ss obtained higher IQ with Es (1 black and 20 whites) of their own race, while on re-test Ss obtained higher IQs with Es of the opposite race. Sattler (1966) reanalyzed the experiment and concluded the results are inconclusive because of methodological deficiencies.

Pasamanick and Knoblock (1955): A white E testing forty 2-year-old black Ss on the Gesell Developmental Examination was claimed to have obtained lower verbal responsiveness scores than presumably would have been obtained by a black E, but no Ss were tested by a black E for comparison.

Forrester and Klaus (1964): Twenty-four black kindergartners obtained nonsignificantly higher Stanford-Binet IQs when tested by a female black E than by a female white E.

La Crosse (1964): A white E obtained significantly lower Stanford-Binet (L-M) retest scores when testing black Ss who had been previously tested by two black Es. The same white E obtained significantly higher retest scores with white Ss previously tested by three white Es.

Pettigrew (1964): White Es (number not reported) are said to obtain fewer correct responses than black Es (number not reported) from Northern blacks given two tests (identification of six famous men and giving synonyms). No statistical tests of significance are reported.

Miller and Phillips (1966): Three black and three white female Es testing black and white children in Head Start in the South resulted in no significant effects, either for race of E, or for the race of E X race of S interaction.

Pelosi (1968): Six black and six white Es tested young adult black males enrolled in a Neighborhood Youth Corps, on the Wechsler Adult Intelligence Scale, the Purdue Pegboard, and the IPAT Culture Fair Test; no significant effects of race of E.

Abramson (1969): Two black and two white female Es gave Peabody Picture Vocabulary Test to Eastern black and white kindergartners and first-graders. No significant E effects for kindergartners, but white Es obtained higher scores from white Ss than from black; and black Es obtained similar scores from both groups.

Lipsitz (1969): Lorge-Thorndike group-administered test-retest by one black and one white E showed no significant race of E or interaction effects in Eastern black and white 4th, 5th, and 6th graders in private schools (unrepresentative samples).

Caldwell and Knight (1970): Stanford-Binet test-retest with one black female E and one white male E produced no significant E effect on 6th grade Southern black children.

Costello (1970): Two white and two black Es giving the Peabody Pciture Vocabulary Test to black preschoolers resulted in no significant race of E effect.

In concluding a detailed critique of this research, Sattler (in press) commented: "The studies reviewed . . . suggest that performance of Negro and white subjects on individually administered intelligence tests is not usually affected by the examiner's race. However, there are still too few studies available to arrive at firm generalizations. Yet, as Sattler (1970, p. 144) pointed out, numerous authorities have stated that difference in racial status is a variable which affects the examiner-examinee relationship. The research cited in this review as well as past research offers no support for this statement."

Sattler (1970, p.144) also points out that "little is known about the effects of the examiners' race on scores obtained on group administered intelligence tests." To examine this matter in terms of available evidence,

Shuey (1966) compared all the reported studies up to 1965 (19 in all) of black IQ in elementary school children in the South, where the group testing was done by a black tester, with the test results obtained on all Southern black school children, the vast majority of whom were tested by white examiners. Shuey concluded:

> "The 2,360 elementary school children tested by Negroes earned a mean IQ of 80.9 as compared with a combined mean of 80.6 earned by more than 30,000 Southern Negro school children, an undetermined but probably a large number of whom were tested by white investigators. The present writer also calculated the combined mean IQ achieved by 1,796 Southern colored high school pupils who were tested by Negro adults. This was 82.9 as compared with a mean of 82.1 secured by nearly 9,000 Southern colored high school students, many of whom were examined by white researchers. From these comparisons it would seem that the intelligence score of a Negro school child or high school pupil has not been adversely affected by the presence of a white tester" (p. 507).

Concerning Shuey's analysis, Dreger and Miller (1968) stated: "Although Shuey, on the basis of finding no apparent differences in school-age Negro children, concludes that such race of E X race of S interaction does not exist, we have offered some reason to think it does exist" (p.25). But the most obvious objection to Shuey's analysis is that there was no control over the samples tested by black and white Es. If for some reason the less intelligent black Ss were more likely to be tested by black Es (as might be the case in most rural Southern schools), the fact that the more intelligent black Ss (more likely in urban schools) tested by white Es did not obtain higher IQs than the Ss tested by black Es might only mean that their performance had been depressed by the presence of a white E. In a proper study pains should be taken to avoid any such biasing factors

in the assignment of white and black Es to white and black Ss

The present study was designed to control such factors and was conducted with large enough samples of both Es and Ss over a sufficient age range and with an adequate variety of mental ability tests as fully to permit the significant appearance of a race of E X race of S interaction. Since statistical significance depends in part upon the sample size, and since the samples in this study are very large, it is much more important to evaluate the actual magnitudes of the examiner effects rather than merely to note their level of statistical significance. All population differences, of course, are significant, and this study is based on virtually the entire elementary school population of a city of more than 100,000 population.

Method

Subjects

The Ss were virtually the total white and black elementary school (kindergarten through 6th grade) population of the Berkeley Unified School District. A total of nearly 9,000 pupils in all classes of 17 schools were tested, with the exclusion only of children in special classes for the retarded, the emotionally disturbed, and the neurologically and physically handicapped. Since the present study focuses upon white-black interaction of race of examiners (E) and race of subjects (Ss), the 11 percent of the school population who are Oriental or other ethnic minorities (about 1 percent) are not included in the analyses. (The total school population involved in this study is approximately 60% white and 40% black.) Also not included are Ss who were absent on the day that a particular test was administered to their class.

Ss' ethnicity was determined from the school records, which included

the parents' statement of the child's race, obtained when the child was enrolled in the Berkeley schools.

Examiners

There were 12 white (10 women and 2 men) and 8 black (6 women and 2 men) Es. All were between 25 and 40 years of ago and all had either B. A. or M. A. degrees in psychology or education. A few were University graduate students in the school psychology program and nearly all of them had teaching credentials and had taught in public schools. They were selected from among some fifty applicants on the basis of qualifications and interviews. They were paid at the daily rate for substitute teachers in the Berkeley schools.

All Es were given copies of the tests and the manuals of instructions for administration to study prior to the three all-day training sessions. These sessions, conducted by three professionals with training and experience in clinical and group testing, aimed to inculcate general principles of test administration as well as specific instructions and practice in the tests to be used. All testing procedures were demonstrated and all Es, working in small groups, had to practice the instructions and procedures in front of the group and the instructor who criticized and "shaped up" each E's performance, in terms of voice, emphasis, pacing, rapport, and general manner of presentation. The importance of strict adherence to the standard instructions and time limits was emphasized repeatedly, and the psychometric rationale for this was thoroughly explained. Es were provided with stopwatches for the timed tests and were taught to operate the tape recorders used in two of the tests. Es were also instructed in filling out a special form at the conclusion of every test session concerning any unusual occurrences (e.g., a fire drill) which might have created nontypical testing conditions.

All Es were observed actually testing in the classroom at least once, early in the testing program, by Dr. Egbert, our testing supervisor, or one of the other professionals on the staff, with the aim of maintaining as much uniformity of testing procedures as possible.

All Es did not administer every one of the different tests used in this study, but every test was administered by white and black Es.

Assignment of Es to Schools and Classes

The assignment of Es to schools and classes was random within race of E. That is, on any given day, one black E was assigned at random to each school until the supply of black Es was used up; the same was done for white Es. Thus, every school received both white and black Es. These random assignments were made on a day-to-day basis, so that all Es had equal chances of testing in all schools. The particular classes to be tested at a given school on a given day also were assigned at random to the white and black Es.

Tests

A variety of quite different tests were used. They were expected possibly to elicit different degrees of sensitivity to examiner effects. There were standard verbal and nonverbal IQ tests, which involved considerable verbal instructions on the part of E, especially in Grades K to 3. There was an untimed developmental perceptual-motor test; a "speed and persistence" test intended to reflect effort and motivation induced by verbal instructions in a test-taking situation; a test of Ss' ability to attend to verbally given directions; and a short-term memory test. Both of these latter two tests involved the presence and supervision of the E as a proctor, but were wholly administered and paced by means of a tape recording to insure

the greatest possible uniformity of administration.

Lorge-Thorndike Intelligence Tests. This is a nationally standardized group-administered test of general intelligence. In the normative sample, which was intended to be representative of the nation's school population, the test has a mean IQ of 100 and a standard deviation of 16. It is generally acknowledged to be one of the best standardized paper-and-pencil tests of general intelligence.

The Manual of the Lorge-Thorndike Test states that the test was designed to measure reasoning ability. It does not test proficiency in specific skills taught in school, although the verbal tests, from Grade 4 and above, depend upon reading ability. The reading level required, however, is intentionally kept considerably below the level of reasoning required for correctly answering the test questions. Thus the test is essentially a test of reasoning and not of reading ability, which is to say that it would have more of its variance in common with nonverbal tests of reasoning ability than with tests of reading per se.

The tests for Grades K-3 do not depend at all upon reading ability but make use exclusively of pictorial items. The tests for Grades 4-8 consist of two parts, Verbal (V) and Nonverbal (NV). They are scored separately and the raw score on each is converted to an IQ, with a normative mean of 100 and SD of 16. The chief advantage of keeping the two scores separate is that the Nonverbal IQ does not overestimate or underestimate the child's general level of intellectual ability because of specific skills or disabilities in reading.

The following forms of the Lorge-Thorndike Intelligence Tests were used:

Level 1, Form B.		Grades K-1
Level 2, Form B.		Grades 2-3
Level 3, Form B	Verbal and Nonverbal.	Grades 4-6.

The "consumable" form of the test was used to obviate separate answer sheets and the added difficulty they may involve for the testees.

Figure Copying Test. This test was developed at the Gesell Institute of Child Study at Yale University as a means for measuring developmental readiness for the traditional school learning tasks of the primary grades (Ilg & Ames, 1967). The test consists of the ten geometric forms arranged in order of difficulty. The child must simply copy them, each on a separate sheet of paper. The test involves no memory factor, since the figure to be copied is before the child at all times. The test is administered without time limit, although most children finish in 10 to 15 minutes. The test is best regarded as a developmental scale of mental ability. It correlates substantially with other IQ tests, but it is considerably less culture-loaded than most usual IQ tests. It is primarily a measure of general cognitive development and not just of perceptual-motor ability. Children taking the test are urged to attempt to copy every figure.

Each of the ten figures is scored on a 3 point scale going from 1 (low) to 3 (high). (A score of zero is given in the rare instance when no attempt has been made to copy a particular figure). A score of 1 is given if an attempt has been made but the child's drawing completely fails to resemble the model. A score of 2 is given if there is fair resemblance to the model--the figure need not be perfect but it must be easily recognizable as the model which the child has attempted to copy. A score of 3 is given for an attempt which duplicates the figure in all its essential characteristics--this is an essentially adult level of performance. Since there are ten figures in all, the possible range of scores goes from 10 to 30 (or 0 to 30 if zeros are counted, but this is rare, since virtually all subjects attempt all ten figures. Scoring reliability, as determined from correlations between

different scores is above .90

The high level of motivation maintained by this test is indicated by the fact that the minimum score obtained in each group at each grade level increases systematically with grade level. This suggests that all children were making an attempt to perform in accordance with the instructions. Another indication that can be seen from the test booklets is that virtually 100 percent of the children in every ethnic group at every grade level attempted to copy every figure. The attempts, even when unsuccessful, usually show considerable effort, as indicated by redrawing the figure, erasures and drawing over the figure repeatedly in order to improve its likeness to the model. It is also noteworthy about this test that normal children are generally not successful in drawing figures beyond their mental age level and that special instructions and coaching on the drawing of these figures hardly improves the child's performance. This test, in other words, is not very susceptible to training, but measures some fundamental aspects of mental development.

Listening-Attention Test. In the Listening-Attention Test the child is presented with an answer sheet containing 100 pairs of digits in sets of 10. The child listens to a tape recording which speaks one digit every two seconds. The child is required to put an X over the one digit in each pair which has been heard on the tape recorder. The purpose of this test is to determine the extent to which the child is able to pay attention to numbers spoken on a tape recorder, to keep his place in the test, and to make the appropriate responses to what he hears from moment to moment. Low scores on this test indicate that the subject is not yet ready to take the Memory for Numbers test which immediately follows it. High scores on the Listening-Attention Test indicate that the subject has the prerequisite skills for taking the digit span (Memory for Numbers) test. The Listening-Attention

Test thus is intended as a means for detecting students who, for whatever reason, are unable to hear and to respond to numbers read over a tape recorder. The test itself makes no demands on the child's memory, but only on his ability for listening, paying attention, and responding appropriately--all prerequisites for the digit memory test that follows.

Memory for Numbers Test. The Memory for Numbers test is a measure of digit span, or more generally, short-term memory. It consists of three parts. Each part consists of six series of digits going from four digits in a series up to nine digits in a series. The digit series are presented on a tape recording on which the digits are spoken clearly by a male voice at the rate of precisely one digit per second. The subjects write down as many digits as they can recall at the conclusion of each series, which is signaled by a "bong." Each part of the test is preceded by a short practice test of three digit series in order to permit the tester to determine whether the child has understood the instructions, etc. The practice test also serves to familiarize the subject with the procedure of each of the subtests. The first subtest is labeled Immediate Recall (I). Here the subject is instructed to recall the series immediately after the last digit has been spoken on the tape recorder. The second subtest consists of Delayed Recall (D). Here the subject is instructed not to write down his response until after ten seconds have elapsed after the last digit has been spoken. The ten-second interval is marked by audible clicks of a metronome and is terminated by the sound of a bong which signals the child to write his response. The Delayed Recall condition invariably results in some retention decrement. The third subtest is the repeated series test, in which the digit series is repeated three times prior to recall; the subject then recalls the series immediately after the last digit in the series has been presented. Again, recall is signaled by a bong. Each repetition of

the series is separated by a tone with a duration of one second. The repeated series almost invariably results in greater recall than the single series. The Memory for Numbers test is very culture fair for children in second grade and beyond and who know their numerals and are capable of listening and paying attention, as indicated by the Listening-Attention Test. The maximum score on any one of the subtests is 39, that is, the sum of the digit series from four through nine. In the present study, the total score is used, i.e., the sum of the three subtest scores.

Speed and Persistence Test (Making X's). The Making X's Test is intended as an assessment of test-taking motivation. It gives an indication of the subject's willingness to comply with instructions in a group testing situation and to mobilize effort in following those instructions for a brief period of time. The test involves no intellectual component, although for young children it probably involves some perceptual-motor skills component, as reflected by increasing mean scores as a function of age between grades 1 to 5. The wide range of individual differences among children at any one grade level would seem to reflect mainly general motivation and test-taking attitudes in a group situation. The test also serves partly as an index of classroom morale, and it can be entered as a moderator variable into correlational analyses with other ability and achievement tests. Children who do very poorly on this test, it can be suspected, are likely not to put out their maximum effort on ability tests given in a group situation and therefore their scores are not likely to reflect their "true" level of ability.

The Making X's Test consists of two parts. On Part I the subject is asked simply to make X's in a series of squares for a period of 90 seconds. In this part the instructions say nothing about speed. They merely instruct the child to make X's. The maximum possible score on

Part I is 150, since there are 150 squares provided in which the child can make X's. After a 2-minute rest period the child turns the page of the test booklet to Part II. Here the child is instructed to show how much better he can perform than he did on Part I and to work as rapidly as possible. The child is again given 90 seconds to make as many X's as he can in the 150 boxes provided. The gain in score from Part I to Part II reflects both a practice effect and an increase in motivation or effort as a result of the motivating instructions, i. e., instruction to work as rapidly as possible.

Results and Discussion

The basic analysis performed on each test at each grade level in which the test was administered is a nested ANOVA, with race of Es nested within race of Ss. Since there were unequal Ns in the four cells of the 2 X 2 (race of E X race of S) design, the main effects for race of E are based on unweighted means for white and black Es; that is to say, in the overall means for white and black Es, equal weights are given to both means despite their unequal Ns. Otherwise the overall mean difference between white and black Es would be partly a function of the number of white and black Ss they had tested, because there is a substantial main effect for race of Ss.

So that the magnitudes of the differences can be readily compared from one grade to another and from one test to another, all differences have been expressed in sigma units. The sigma in every case is the standard deviation of test scores within groups, i.e., the standard deviation excluding variance due to race of E, race of Ss, and their interaction.

Table 1 shows the Ns of Es and Ss for each of the tests at each grade.

- - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - -

Lorge-Thorndike IQ

Tables 2 and 3 show the results for Lorge-Thorndike Nonverbal and Verbal IQs, respectively. For Nonverbal IQ, the main effect of race of

- - - - - - - - - - - - - - - - - - -

Insert Tables 2 and 3 about here

- - - - - - - - - - - - - - - - - -

E, as we can see in the first column, is very small; only at Grades 1 and 2 is the difference significant, and it amounts to less than a fifth of a standard deviation, or less than 3 IQ points. (Negative numbers always indicate that the black mean exceeded the white.) The overall mean difference between white and black Es (shown in the last two rows of the first column) is very small and nonsignificant even for these very large samples. The unweighted mean ($\bar{X}$) here is the simple arithmetic average of the means of every grade; the weighted mean is the average of the means of every grade, each weighted by the total number of Ss on which the mean is based. Since the Ns are usually similar from one grade to another, the weighted and unweighted means do not differ appreciably.

Columns 2 and 3 show the mean differences between white and black Es within each racial group of Ss. Again, these differences are very small, and overall they are nonsignificant for the Nonverbal test. Column 4 shows

Table 1

Number of Examiners (in Italics) and Number of Subjects (in Roman type)

of Each Race (W = White, B = Black) in Each Grade for Each Test

Grade		K		1		2		3		4		5		6		Total Ss	
Tests	Es	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B
Arge-Thorndike Nonverbal IQ	W	*8* 468	*8* 333	*7* 447	*6* 294	*7* 394	*6* 350	*5* 308	*5* 247	*7* 361	*7* 248	*7* 298	*7* 184	*8* 366	*8* 237	2642	1893
	B	*3* 132	*3* 58	*3* 194	*2* 107	*4* 297	*4* 139	*3* 273	*2* 148	*5* 226	*4* 165	*6* 254	*6* 205	*4* 248	*4* 166	1624	988
Arge-Thorndike Verbal IQ	W									*6* 363	*7* 270	*8* 393	*7* 89	*8* 434	*7* 208	1190	567
	B									*5* 202	*6* 100	*4* 161	*5* 302	*4* 171	*5* 151	534	553
Figure Copying	W	*8* 291	*8* 268	*9* 363	*9* 289	*11* 510	*11* 337	*11* 386	*11* 222	*8* 226	*8* 211					1776	1327
	B	*7* 340	*6* 211	*7* 321	*7* 196	*6* 174	*6* 153	*7* 239	*7* 157	*7* 257	*7* 207					1331	924
Speed and Persistence	W			*11* 507	*11* 325	*10* 440	*10* 309	*10* 456	*10* 261	*9* 378	*10* 281	*11* 429	*12* 308	*10* 446	*10* 245	2656	1729
	B			*7* 195	*8* 172	*6* 169	*7* 187	*5* 157	*5* 150	*6* 247	*7* 150	*7* 167	*7* 111	*6* 179	*6* 144	1114	914
Listening-Attention and Memory	W					*4* 471	*4* 347	*4* 358	*4* 260	*4* 330	*4* 300	*5* 337	*5* 289	*5* 374	*5* 270	1870	1466
	B					*2* 178	*2* 114	*2* 219	*2* 133	*2* 221	*2* 85	*2* 210	*2* 103	*2* 197	*2* 99	1025	534

Table 2

Lorge-Thorndike Nonverbal IQ:

Mean Differences in Sigma Units

Grade	Mean W-B E Difference	Mean W-B E Difference		Mean W-B S Difference[a]	Between Es Within Groups[b]
		White Ss	Black Ss		
K	.062	.062	.062	1.19	.252
1	-.199**	-.219**	-.180	1.37	.466
2	-.166**	-.192*	-.140	1.31	.523
3	-.011	.089	-.112	1.33	.369
4	.023	.295**	-.249*	1.63	.392
5	.080	.105	.056	1.75	.696
6	.065	.159	-.029	1.73	.517
Unweighted $\bar{X}$	-.021	.043	-.085	1.47	.459
Weighted $\bar{X}$	-.026	.047	-.088	1.47	.460

[a]All differences significant beyond .01.

[b]Not tested for significance.

*Significant at $p < .05$.

**Significant at $p < .01$.

Table 3

Lorge-Thorndike Verbal IQ:

Mean Differences in Sigma Units

Grade	Mean W-B E Difference	Mean W-B E Difference		Mean W-B S Difference[a]	Between Es Within Groups[b]
		White Ss	Black Ss		
4	-.003	-.013	.007	1.59	.680
5	.404**	.371**	.437**	1.60	.415
6	.296**	.422**	.170	1.95	.710
Unweighted $\bar{X}$	.232**	.260**	.205**	1.71	.602
Weighted $\bar{X}$	.233**	.263**	.209**	1.71	.602

[a]All differences significant beyond .01.

[b]Not tested for significance.

**Significant at $p < .01$.

the race difference between Ss, against which one can compare the magnitudes of the differences shown in the other columns.

The last column shows the variation among Es within groups, that is, variation among Es not attributable to race of E or race of Ss or the interaction of these variables. This variation among Es is expressed as the standard deviation of the means of Es within groups divided by the standard deviation of Ss within groups. It should be noted, however, that some appreciable part of the variation among Es reflects differences between schools and classrooms, which inevitably results from the random assignment of a relatively small number of Es to a diversity of schools and classes. The interschool and interclass variations do not have a chance to "average out" over Es under the conditions of the present study. The between E variation allows no meaningful test of statistical significance but is presented here merely as a basis for comparing and evaluating the magnitudes of the other differences.

These general comments serve as well for Tables 6 through 10.

On the Verbal IQ (Table 3) we see that the race of E differences, both for the main effect and within groups, show the white Es obtaining slightly higher scores by about one-fourth to one-fifth of a sigma. But the net effect of race of E is an overall mean difference between white and black Ss of only about .05 sigma, or less than one IQ point, as compared with the overall 1.71 sigma difference between the white and black Ss means.

Individually Administered IQ Tests. To determine if individual administration of the IQ tests would result in significant race of E by race of S interaction, a number of Ss were tested individually on the Lorge-Thorndike Tests. At the time that a group test was administered to a class, one child in the class was taken at random from the class to be given the same test individually, either in a private room in the school

or in a testing van parked on the school grounds. The child was selected by random numbers from the class roll, irrespective of race or sex, prior to the time of testing; in the case of absentees, there was always an alternate who had been selected by the same random procedure. Every single class in the school system was thus represented by one child selected at random for an individual test. Equal numbers of white and black Es were assigned at random as individual testers.

The essential results are shown in Tables 4 and 5. A two-way ANOVA

- - - - - - - - - - - - - - - - -

Insert Tables 4 and 5 about here

- - - - - - - - - - - - - - - - -

was done at each grade. The race of S main effect was significant beyond the .01 level in every grade. Tables 4 and 5 give the exact p values of F in the ANOVAs for the race of E main effect and for the race of E X race of S interaction. With one exception (Verbal IQ at Grade 5) the p values of the interaction fall far short of significance, and in general the race of E has no appreciable or systematic effect on the individually administered tests.

Figure Copying Test

Results for the Figure Copying Test are shown in Table 6. The race

- - - - - - - - - - - - - - - -

Insert Table 6 about here

- - - - - - - - - - - - - - - -

of E effects can be seen to be quite small and unsystematic.

Table 4

Lorge-Thorndike Nonverbal IQ: Individual Tests

Grades		K Ss		1 Ss		2 Ss		3 Ss		4 Ss		5 Ss		6 Ss		All Grades Ss	
Es		W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B
N	W	18	7	24	6	13	15	12	8	16	13	16	12	14	6	113	67
	B	7	17	7	16	17	9	11	12	6	7	9	13	10	8	67	82
IQ:$\bar{X}$	W	108.7	86.4	108.3	87.5	106.1	88.9	102.4	81.7	124.7	102.8	116.0	98.3	124.6	95.8	112.9	92.7
	B	104.0	90.2	106.9	92.6	110.1	90.4	116.5	102.0	113.3	105.9	114.0	92.0	114.1	92.9	111.6	94.3
P value																	
Race of E		.92		.74		.56		.00		.46		.60		.03		.95	
Interaction E x S		.37		.46		.81		.54		.18		.75		.28		.11	

Table 5

Lorge-Thorndike Verbal IQ: Individual Tests

Grades		4 Ss		5 Ss		6 Ss		All Grades Ss	
Es		W	B	W	B	W	B	W	B
N	W	21	8	15	16	13	7	49	31
	B	5	10	7	7	5	11	17	28
IQ : $\bar{X}$	W	114.5	92.6	123.1	83.9	117.3	87.9	117.9	87.0
	B	118.2	94.5	103.1	98.6	110.2	85.9	109.6	92.1
P value Race of E		.52		.58		.46		.59	
Interaction E X S		.83		.00		.66		.00	

Table 6

Figure Copying Test:

Mean Differences in Sigma Units

Grade	Mean W-B E Difference	Mean W-B E Difference		Mean W-B S Difference[a]	Between Es Within Groups[b]
		White Ss	Black Ss		
K	.002	-.015	.019	1.00	.317
1	-.076	-.009**	-.144	.95	.343
2	-.048	-.079	-.017	.85	.420
3	.270**	.354**	.185	.87	.539
4	.078	.237**	-.081	.99	.521
Unweighted $\bar{X}$	.045	.098**	-.008	.93	.428
Weighted $\bar{X}$	.037	.085*	-.015	.93	.421

[a]All differences significant beyond .01.

[b]Not tested for significance.

*Significant at $p < .05$.

**Significant at $p < .01$.

Speed and Persistence Test (Making Xs).

It is interesting that this test, which was devised to reflect Ss' attitude and effort in a test situation and to be sensitive to motivating instructions, does in fact show far larger E effects than any of the other tests used in this study, differences amounting to half a standard deviation or more. It also shows by far the smallest overall racial difference between Ss of any of the tests. The consistently significant race of E effects uniformly favor the white Es. The neutral and motivating instructions do

- - - - - - - - - - - - - - - - - -

Insert Tables 7 and 8 about here

- - - - - - - - - - - - - - - - - -

not appear to produce any differnces with respect to the race variables. However, it should be remarked that performances under the motivating instructions is significantly higher for all groups than under the neutral instructions.

Listening-Attention and Memory for Numbers

These two tests were expected to show the smallest E effects, since their administration was wholly by means of a tape recording and involved the Es only as proctors and distributors of test forms. This expectation is borne out by the results shown in Tables 9 and 10.

- - - - - - - - - - - - - - - - - -

Insert Tables 9 and 10 about here

- - - - - - - - - - - - - - - - - -

Table 7

Speed and Persistence Test--First Try (Neutral Instructions):

Mean Differences in Sigma Units

Grade	Mean W-B E Difference	Mean W-B E Difference		Mean W-B S Difference	Between Es Within Groups[a]
		White Ss	Black Ss		
1	.178**	.287**	.069	.56**	.812
2	.508**	.370**	.693**	-.09	.764
3	.542**	.588**	.496**	-.20**	.854
4	1.147**	1.185**	1.109**	-.44**	1.043
5	.550**	.650**	.450**	.10	1.062
6	.500**	.638**	.362**	.17**	.982
Unweighted $\bar{X}$	.571**	.620**	.530**	.02	.919
Weighted $\bar{X}$	.562**	.614**	.526**	.03	.915

[a]Not tested for significance.

**Significant at $p < .01$.

Table 8

Speed and Persistence Test--Second Try (Motivating Instructions):

Mean Differences in Sigma Units

Grade	Mean W-B E Difference	Mean W-B E Difference		Mean W-B S Difference	Between Es Within Groups[a]
		White Ss	Black Ss		
1	.265**	.460**	.070	.53**	.793
2	.617**	.641**	.594**	.07	.727
3	.685**	1.013**	.357**	-.18**	.818
4	1.019**	1.263**	.775**	-.55**	1.086
5	.387**	.621**	.125	-.03	1.093
6	.477**	.630**	.324**	-.03	1.045
Unweighted $\bar{X}$	.575**	.771**	.374**	-.03	.927
Weighted $\bar{X}$	.570**	.766**	.374**	-.02	.921

[a]Not tested for significance.

**Significant at $p < .01$.

Table 9

Listening-Attention Test:

Mean Differences in Sigma Units

Grade	Mean W-B E Difference	Mean W-B E Difference		Mean W-B S Difference[a]	Between Es Within Groups[b]
		White Ss	Black Ss		
2	-.055	.020	-.131	.23	.119
3	-.288**	-.039	.022	.36	.112
4	-.121	.036	-.279**	.32	.141
5	.059	.046	.071	.18	.461
6	.145*	.098	.192*	.19	.203
Unweighted $\bar{X}$	-.052	.032	-.025	.25	.207
Weighted $\bar{X}$	-.053	.031	-.030	.25	.203

[a] All differences significant beyond .01.

[b] Not tested for significance.

* Significant at $p < .05$.

** Significant at $p < .01$.

Table 10

Memory for Numbers Test:

Mean Differences in Sigma Units

Grade	Mean W-B E Difference	Mean W-B E Difference		Mean W-B S Difference[a]	Between Es Within Groups[b]
		White Ss	Black Ss		
2	-.070	-.229**	-.125	.61	.114
3	.161*	.118	.204	.58	.201
4	.062	-.055	.180	.59	.215
5	.105	.063	.147	.67	.254
6	-.002	-.143	.139	.72	.236
Unweighted $\bar{X}$	.051	-.049	.109	.63	.204
Weighted $\bar{X}$	.048	-.055	.100	.63	.201

[a]All differences significant beyond .01.

[b]Not tested for significance.

**Significant at $p < .01$.

Conclusion

Readers must evaluate the magnitudes of the race of E effects not so much in terms of their level of statistical significance but in relation to the magnitudes of the other sources of variance in test scores and in relation to the size of differences in mental test scores that are of practical or theoretical consequence in any particular context. From the present results for the tests of cognitive ability (i.e., excepting the Speed and Persistence Test), it seems safe to conclude that for all practical purposes the race of the examiner is of negligible consequence in the testing of white and black school children.

References

Abramson, T. The influence of examiner race on first-grade and kindergarten subjects' Peabody Picture Vocabulary Test scores. Journal of Educational Measurement, 1969, 6, 241-46.

Caldwell, M. B., & Knight, D. The effect of Negro and white examiners on Negro intelligence test performance. Journal of Negro Education, 1970, 39, 177-79.

Canady, H. G. The effect of "rapport" on the IQ: A new approach to the problem of racial psychology. Journal of Negro Education, 1936, 5, 209-219.

Costello, J. Effects of pretesting and examiner characteristics on test performance of young disadvantaged children. Proceedings of the 78th Annual Convention of the American Psychological Association, 1970, 5, 309-10.

Dreger, R. M., & Miller, K. S. Comparative psychological studies of Negroes and white in the United States: 1959-1965. Psychological Bulletin Monograph Supplement, 1968, 70, No. 3, Part 2.

Forrester, B. J., & Klaus, R. A. The effect of race of the examiner on intelligence test scores of Negro kindergarten children. Peabody Papers in Human Development, 1964, 2, 1-7.

Ilg, Frances L. & Ames, Louise Bates. School readiness: Behavior tests used at the Gesell Institute. New York: Harper & Row, 1964.

La Crosse, J. E. Examiner reliability on the Stanford-Binet Intelligence Scale (Form L-M) in a design employing white and Negro examiners and subjects. Unpublished master's thesis, University of North Carolina, 1964.

Lipsitz, S. Effect of the race of the examiner on results of intelligence test performance of Negro and White children. Unpublished master's thesis, Long Island University, 1969.

Miller, J. O., & Phillips, J. A preliminary evaluation of the Head Start and other metropolitan Nashville kindergartens. Unpublished manuscript, Demonstration and Research Center for Early Education, George Peabody College for Teachers, Nashville, 1966.

Pasamanick, B., & Knobloch, H. Early language behavior in Negro children and the testing of intelligence. Journal of Abnormal and Social Psychology, 1955, 50, 401-402.

Pelosi, J. W. A study of the effects of examiner race, sex, and style on test responses of Negro examinees. Unpublished doctoral dissertation, Syracuse University, 1968.

Pettigrew, T. F. A profile of the Negro American. Princeton: D. Van Nostrand, 1964.

Sattler, J. M. Statistical reanalysis of Canady's "The effect of 'rapport' on the IQ: A new approach to the problem of racial psychology." Psychological Reports, 1966, 19, 1203-6.

Sattler, J. M., & Theye, F. Procedural, situational, and interpersonal variables in individual intelligence testing. Psychological Bulletin, 1967, 68, 347-360.

Sattler, J. M. Racial "experimenter effects" in experimentation, testing, interviewing, and psychotherapy. Psychological Bulletin, 1970, 73, 137-160.

Sattler, J. M. Racial experimenter effects. In K. S. Miller & R. M. Dreger (Eds.), Comparative physiological, psychological, and sociological studies of Negroes and Whites in the United States: 1966-1970. Seminar Press, in press.

Shuey, Audrey M. The testing of Negro intelligence. (2nd ed.) New York: Social Science Press, 1966.